

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comogen Ltd.

OM protein - protein search, using SW model

Run on: March 1, 2001, 15:49:51 ; Search time 140.11 Seconds  
(without alignments)  
5.383 Million cell updates/sec

Title: US-09-331-631a-8\_COPY\_120\_161

Sequence: 1 SORFOECQCHQCHQOEORPE.....QOCVRECKEYQENPMRGR 42

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/pdata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/pdata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/pdata/2/1aa/6.COMB.pep:\*  
4: /cgn2\_6/pdata/2/1aa/PCTRUS.COMB.pep:\*  
5: /cgn2\_6/pdata/2/1aa/backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	225	91.8	587	1	US-07-955-905A-23
2	111.5	45.5	566	1	US-07-955-905A-2
3	111.5	45.5	566	1	US-07-955-905A-22
4	66	26.9	816	2	US-08-267-803B-9
5	64	26.1	1898	1	US-08-056-200-94
6	64	26.1	1898	2	US-08-800-644-94
7	63	25.7	184	1	US-08-453-924-3
8	61.5	25.1	34	2	US-08-822-561-1
9	61	24.9	678	4	PCF-US93-03027-3
10	58	23.7	303	1	US-08-185-432-5
11	58	23.7	737	1	US-08-185-432-2
12	58	23.7	737	1	US-08-185-432-4
13	58	23.7	2703	1	US-08-185-432-19
14	55	22.4	397	1	US-08-647-928-8
15	54.5	22.2	571	1	US-07-955-905A-25
16	53	21.6	131	3	US-09-067-782A-4
17	53	21.6	351	3	US-09-067-782A-2
18	53	21.6	1162	3	US-08-728-323A-2
19	52.5	21.4	205	1	US-08-428-415-5
20	52.5	21.4	205	1	US-08-379-685-5
21	52.5	21.4	205	2	US-08-854-029-5
22	52.5	21.4	523	1	US-08-073-383-2
23	52.5	21.4	523	1	US-08-428-415-2
24	52.5	21.4	523	1	US-08-379-685-2
25	52.5	21.4	523	2	US-08-854-029-2
26	52.5	21.4	523	4	PCF-US94-06365-2
27	52.5	21.4	591	3	US-08-965-903B-2
28	52	21.2	69	2	US-08-690-011A-7

29	52	21.2	93	2	US-08-690-011A-9	Sequence 9, Appl
30	52	21.2	245	1	US-07-876-284-2	Sequence 2, Appl
31	52	21.2	245	1	US-08-276-151-9	Sequence 9, Appl
32	52	21.2	246	1	US-08-276-151-7	Sequence 7, Appl
33	52	21.2	716	2	US-08-766-982-1	Sequence 1, Appl
34	52	21.2	3144	1	US-08-246-982A-6	Sequence 6, Appl
35	52	21.2	3144	1	US-08-453-265-6	Sequence 42, Appl
36	52	21.2	3144	2	US-08-457-273B-42	Sequence 21, Appl
37	52	21.0	3144	3	US-08-556-419-21	Sequence 3, Appl
38	51.5	21.0	513	3	US-09-100-193-3	Sequence 2, Appl
39	51	20.8	138	2	US-08-684-101-2	Sequence 8, Appl
40	51	20.8	215	2	US-08-659-251-8	Sequence 2, Appl
41	51	20.8	215	4	PCF-US96-11445-8	Sequence 8, Appl
42	51	20.8	360	2	US-08-531-927B-2	Sequence 4, Appl
43	51	20.8	361	1	US-08-415-751-4	Sequence 4, Appl
44	51	20.8	2616	5	5206163-3	Patent No. 5206163
45	50.5	20.6	376	5	5180810-1	Patent No. 5180810

## ALIGNMENTS

```
RESULT 1
US-07-955-905A-23
; Sequence 23, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
; PRECURSOR
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 587 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..587
; OTHER INFORMATION: /note="Vicilin from G. hirsutum"
US-07-955-905A-23

Query Match 91.8%; Score 225; DB 1; Length 587;
Best Local Similarity 90.5%; Pred. No. 7e-19;
Matches 38; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SORFOECQCHQCHQOEORPEKQOCVRECKEYQENPMRGR 42
DB 119 SORFOECQCHQCHQOEORPEKQOCVRECKEYQENPMRGR 160

RESULT 2
US-07-955-905A-2
; Sequence 2, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
; PRECURSOR
; NUMBER OF SEQUENCES: 28
```











```

1      TITLE OF INVENTION:  PRECURSOR
2      NUMBER OF SEQUENCES:  28
3      COMPUTER READABLE FORM:
4      MEDIUM TYPE:  Floppy disk
5      COMPUTER:  IBM PC compatible
6      OPERATING SYSTEM:  PC-DOS/MS-DOS
7      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
8      CURRENT APPLICATION DATA:
9      APPLICATION NUMBER:  US/07/955,905A
10     FILING DATE:  21-JAN-1993
11     CLASSIFICATION:  435
12     INFORMATION FOR SEQ ID NO:  25:
13     SEQUENCE CHARACTERISTICS:
14     LENGTH:  571 amino acids
15     TYPE:  amino acid
16     TOPOLOGY:  Linear
17     MOLECULE TYPE:  Protein
18     ORIGINAL SOURCE:
19     ORGANISM:  Pisum sativum
20     FEATURE:
21     NAME/KEY:  Protein
22     LOCATION:  1..571
23     OTHER INFORMATION:  /note= "Conviciilin from P. sativum"
24
25     US-07-955-905A-25

```

Query Match	22.2%	Score	54.5	DB	1	Length	571
Best Local Similarity	31.0%	Pred. No.	46				
Matches	13	Conservative	12	Mismatches	12	Indels	5
						Gaps	1

```
Oy      6 QECQCHHOEORPEKKOCVRECREKY----ENPMKEER 42
          :|:::|:::| | | :
Db      70 EEKQRYRVGREKKEQKEYQGPRRMEEFDEQEVEEMRSQ 111
```

```
Search completed: March 1, 2001, 15:49:53
Job time: 398 sec
```

